

Quasar Probabilities and Redshifts from *WISE* mid-IR through *GALEX* UV Photometry

M.A. DiPompeo¹, J. Bovy^{2*}, A.D. Myers¹, D. Lang³

¹ *University of Wyoming, Dept. of Physics and Astronomy 3905, 1000 E. University, Laramie, WY 82071, USA*

² *Institute for Advanced Study, Einstein Drive, Princeton, NJ 08450, USA*

³ *McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*

Accepted 2015 July 10; Received 2015 July 9; in original form 2014 December 02

ABSTRACT

Extreme deconvolution (XD) of broad-band photometric data can both separate stars from quasars and generate probability density functions for quasar redshifts, while incorporating flux uncertainties and missing data. Mid-infrared photometric colors are now widely used to identify hot dust intrinsic to quasars, and the release of all-sky *WISE* data has led to a dramatic increase in the number of IR-selected quasars. Using forced-photometry on public *WISE* data at the locations of SDSS point sources, we incorporate this all-sky data into the training of the *XDQSOz* models originally developed to select quasars from optical photometry. The combination of *WISE* and SDSS information is far more powerful than SDSS alone, particularly at $z > 2$. The use of SDSS+*WISE* photometry is comparable to the use of SDSS+ultraviolet+near-IR data. We release a new public catalogue of 5,537,436 (total; 3,874,639 weighted by probability) potential quasars with probability $P_{\text{QSO}} > 0.2$. The catalogue includes redshift probabilities for all objects. We also release an updated version of the publicly available set of codes to calculate quasar and redshift probabilities for various combinations of data. Finally, we demonstrate that this method of selecting quasars using *WISE* data is both more complete and efficient than simple *WISE* color-cuts, especially at high redshift. Our fits verify that above $z \sim 3$ *WISE* colors become bluer than the standard cuts applied to select quasars. Currently, the analysis is limited to quasars with optical counterparts, and thus cannot be used to find highly obscured quasars that *WISE* color-cuts identify in significant numbers.

Key words: methods: data analysis; catalogues; galaxies: active; galaxies: distances and redshifts; galaxies: photometry; (galaxies:) quasars: general

1 INTRODUCTION

As newer and larger imaging surveys are conducted over more area and frequency ranges, photometric classification of quasars using as much available information as possible is becoming increasingly important. Spectroscopic follow-up of complete samples will pose greater and greater technical challenges as the depth of imaging surveys increases. Studies of photometric quasar samples have already shed light on the full quasar population (Scranton et al. 2005; Giannantonio et al. 2006, 2008; Myers et al. 2006, 2007a,b; Hickox et al. 2011; Donoso et al. 2014; DiPompeo et al. 2014). Not only is classification important, but photometric redshift probability density functions (PDFs) are also useful for many applications (e.g. Myers et al. 2006, 2007a). The ability to generate photo-*zs* for quasars has improved with deeper and more precise multi-filter photometry (Richards et al. 2001a,b; Budavári et al. 2001; Bovy et al. 2012).

The problem of photometric quasar classification is well studied, and often involves breaking up samples into redshift bins that allow the classification schemes to also serve as broad redshift indicators (e.g. Richards et al. 2004, 2009a,b; Suchkov, Hanisch & Margon 2005; Ball et al. 2007, 2008). Bovy et al. (2011) developed one of the most successful current quasar classification techniques (*XDQSO*), by using a large number of Gaussians to model the flux space of quasars and stars/unresolved galaxies (incorporating missing and/or noisy data using *extreme deconvolution*, *XD*; Bovy, Hogg & Roweis 2011) to assign quasar probabilities in broad redshift bins. This method was successfully applied in selecting quasars (primarily at $z > 2$) for the Baryon Oscillation Spectroscopic Survey (BOSS; Ross et al. 2012; Dawson et al. 2013). Bovy et al. (2012) extended this method to incorporate redshift information into the model (*XDQSOz*) such that it can be integrated analytically to provide quasar probabilities over arbitrary redshift ranges and generate photometric redshift PDFs. Bovy et al. (2012) also incorporated ultraviolet (UV) and near-infrared (NIR) forced-photometry at

* Hubble Fellow, John N. Bahcall Fellow

SDSS source positions into the model, which improved quasar classification and essentially broke all redshift degeneracies evident when using photometry from the Sloan Digital Sky Survey (SDSS; York et al. 2000) *ugriz* filters (Fukugita et al. 1996) alone.

While Bovy et al. (2012) illustrated the power of NIR photometry in the *XDQSOz* method, its use is limited because of the shallow depths and relatively smaller areas of current NIR surveys (see section 2.2.4). It is now well established that the mid-IR is an efficient identifier of quasars (Stern et al. 2005; Hickox et al. 2007), and the public all-sky data from the *Wide-Field Infrared Survey Explorer* (*WISE*; Wright et al. 2010) allows the use of this method to identify very large samples. With no additional information, over redshifts from $0 < z < 3$, *WISE* colors can select luminous quasars at completeness levels near 90% (at conservative flux limits of $m \lesssim 15$; Stern et al. 2012; Assef et al. 2013). These techniques have been used to classify the first large samples of obscured quasars (Mateos et al. 2012, 2013) and begin to statistically compare obscured and unobscured quasar populations (Donoso et al. 2014; DiPompeo et al. 2014).

In this paper, we incorporate the mid-IR data of *WISE* in the *XDQSOz* model of the relative-flux-redshift density, and show that these data dramatically increase the power for *XDQSOz* to identify quasars, especially at high redshift, as well as improve photometric redshift estimation. We also present a new catalogue of quasar probabilities for point sources in SDSS DR8, including a potential quasar catalogue for objects with probabilities above 0.2 that includes redshift PDFs. We caution that this catalogue is probabilistic in nature, and not suitable as a statistical sample (at least not in its entirety — statistical subsamples can be selected from the full catalogue). Section 3.3 discusses this in more detail.

2 METHODS & DATA

2.1 Photometric classification and redshift estimation

Details of the method used to calculate quasar/star probabilities for objects and redshift PDFs are presented fully in Bovy et al. (2011) and Bovy et al. (2012). We refer the reader there for complete details, with a brief summary of the general considerations provided here.

The important factor for photometrically classifying quasars and estimating redshifts is the joint probability of an object's fluxes, redshift, and the possibility that it is a quasar: $p(\text{flux}, z, \text{quasar})$. This can be written in many different ways depending on the approach, but *XDQSOz* is based on:

$$p(\text{fluxes}, z, \text{quasar}) = p(\text{fluxes}, z | \text{quasar}) P(\text{quasar}). \quad (1)$$

Under the assumption that an object is a quasar, photometric redshifts are given by:

$$p(z | \text{fluxes}, \text{quasar}) = \frac{p(\text{fluxes}, z, \text{quasar})}{p(\text{fluxes}, \text{quasar})}. \quad (2)$$

To determine if an object is a quasar in a given redshift range, we integrate the joint probability over redshift:

$$P(\text{quasar in } \Delta z | \text{fluxes}) = \int_{\Delta z} dz p(\text{quasar}, z | \text{fluxes}) \quad (3)$$

$$= \int_{\Delta z} dz \frac{p(\text{quasar}, z, \text{fluxes})}{p(\text{fluxes})} \quad (4)$$

where

$$p(\text{fluxes}) = p(\text{fluxes}, \text{quasar}) + p(\text{fluxes}, \text{not a quasar}) \quad (5)$$

and $p(\text{fluxes}, \text{not a quasar})$ is found by modeling the fluxes of non-quasars (the stellar sample described below). The probability that an object is a quasar at any z is calculated with $\Delta z = [0, \infty]$. Given that the quasar training set is in the range $0.3 < z < 5.5$, in practice these are the redshifts that should be considered and are the limits applied when we refer to a quasar at “any redshift”.

2.2 Training data

The training sets used are essentially identical to those of Bovy et al. (2011) and Bovy et al. (2012), with the addition of *WISE* data described in section 2.2.5, and so we only provide a brief summary of these, referring the reader to the original papers for full details.

2.2.1 Sloan Digital Sky Survey optical data

The SDSS imaged $\sim 10,000 \text{ deg}^2$ of the northern and southern Galactic sky in *u*, *g*, *r*, *i*, and *z* filters. SDSS-III extended this by $\sim 2500 \text{ deg}^2$ in the southern Galactic cap (Eisenstein et al. 2011). In addition, the SDSS has taken follow-up spectroscopy of millions of sources, which have generated several catalogues of various source subclasses. All of the optical data used to train the *XDQSOz* algorithms are based on SDSS imaging and spectroscopy.

2.2.2 Quasars and stars

Our goal is to classify point-like objects as either stars or quasars. This of course assumes that all resolved sources are galaxies, automatically removing them from the samples, or conversely that all quasars are unresolved. While not strictly true, this is largely handled by limiting the redshifts probed to $z \geq 0.3$. Limiting to unresolved sources will discard a significant fraction of optically faint obscured (“type 2”) quasars, but should not strongly impact optically bright unobscured (“type 1”) sources.

The quasar training data consists of 103,601 spectroscopically confirmed quasars from the SDSS data release 7 (DR7) quasar catalogue (Schneider et al. 2010) with $z \geq 0.3$. While the training is performed in bins of SDSS *i*-band magnitude, all of the quasars are included in each bin in order to have large enough numbers to properly constrain the fits. Thus the quasar flux in each filter is rescaled based on its *i*-band magnitude and redshift, using an apparent-magnitude-dependent redshift prior. The prior is obtained by integrating a model of the quasar luminosity function (Hopkins, Richards & Hernquist 2007)¹ over the appropriate apparent magnitude bin (see Figure 1 of Bovy et al. 2012, and section 2.2.2 for more details). This is used to re-weight the quasars in each *i*-band bin so that the weighted histogram of the quasars reproduces the prediction of the luminosity function. This method also works under the assumption that quasar colors are independent of their absolute magnitude. While there are known correlations between quasar properties and luminosity (e.g. Baldwin 1977; Yip et al. 2004), these effects are small compared to the intrinsic scatter and are generally washed out in broadband colors.

To model the flux space of stars, we use a large set of point sources from the co-added photometry of SDSS Stripe 82

¹ Preliminary work with BOSS quasars suggests that this luminosity function may be incorrect above $z \sim 2.5$ (Ross, private communication). At present, this is still the best choice, but it is possible future updates may be needed.

(Abazajian et al. 2009) that are selected to have low variability, because essentially all quasars are variable and most stars are not (Kirkpatrick et al. 2011). This identifies a stellar training set of 701,215 objects. Of those with available spectra (23,540), only 221 are quasars, indicating that the contamination is extremely low. Additionally, as unresolved galaxies are also not highly variable, these objects are included as part of the stellar training set and thus accounted for when the quasar probability is calculated.

2.2.3 Galaxy Evolution Explorer UV data

We utilize data from the all-sky *Galaxy Evolution Explorer* (GALEX; Martin et al. 2005) in the near- and far-UV (NUV and FUV). Rather than using GALEX catalogue fluxes, we use values force-photometered from GALEX images at SDSS source positions (Aihara et al. 2011). This allows us to obtain low signal-to-noise PSF fluxes of objects that are not detected by GALEX, and thus not included in the GALEX catalogues. One update in this work as compared to Bovy et al. (2012) is that the GALEX forced photometry was not complete at the time of the original analysis, and so UV data for more objects are included here — 81,011 quasars have NUV data, and 68,523 have FUV data, with the numbers increasing slightly for those that only have upper limits. This additional UV data in the training set improves XDQSOz slightly on its own (see section 3.1). The distributions of the UV signal-to-noise ratio (SNR) of these sources are shown in the middle panel of Figure 1.

2.2.4 UKIRT Infrared Deep Sky Survey NIR data

The UKIRT Infrared Deep Sky Survey (UKIDSS) covers ~ 4000 deg² of the SDSS footprint in the *Y*, *J*, *H*, and *K* NIR filters. We make use of this imaging, again force-photometered at SDSS source positions. We find that $\sim 26,000$ objects have complementary fluxes in all of these filters, again with the number increasing when only those with upper limits are considered. The numbers of objects with flux measurements per band are 26,487 (*Y*); 26,450 (*J*); 26,486 (*H*); 26,561 (*K*). Again, these numbers are slightly updated from Bovy et al. (2012), though the differences are quite small. The distributions of SNR values for the UKIDSS imaging are shown in the bottom panel of Figure 1.

2.2.5 Wide-Field Infrared Survey Explorer MIR data

WISE has mapped the entire sky multiple times, in four bands centered at 3.4, 4.6, 12, and 22 μm (*W1*, *W2*, *W3*, and *W4* respectively). The 5σ limit in each band is at least 0.08, 0.11, 1, and 6 mJy, respectively, and improves toward the ecliptic poles where the observing strategy leads to deeper imaging. The angular resolutions in each band are 6.1, 6.4, 6.5, and 12 arc sec, respectively. WISE has released two full-sky source catalogs — however, again in order to use the full power of the XD method and incorporate information on sources below the flux limit of the survey, we make use of forced photometry of the WISE All-Sky Release imaging at SDSS positions (to be included in a future SDSS release; see Lang 2014; Lang, Hogg & Schlegel 2014).

We utilize the two most sensitive bands in WISE, *W1* and *W2*, also the two bands most efficient for selecting quasars via their IR colors. We find 103,050 quasars with *W1* fluxes and 103,019 with *W2* fluxes. The SNR distributions for these sources are shown in Figure 1. Not only does WISE have the most data for the training set, it is also the deepest.

One possible complication with using WISE forced-photometry is that it is performed using source positions from DR12 imaging while this work uses DR8 imaging. In addition to the known astrometry error that was corrected with DR9 (especially above Dec $\sim 41^\circ$; Ahn et al. 2012), some objects have updated “primary” imaging from DR8 to DR12, causing them to shift positions or change primary run number. Our catalogue is built by running XDQSOz on all available data for an object, matching objects in the forced-photometry catalogues by run. If an object’s primary SDSS observation changes runs, we may not find its corresponding WISE data. However, testing reveals that these issues affect a very small fraction of objects — over 99.9% of DR8 primary point sources match to force-photometered WISE sources in the same runs in DR12. Of the unmatched objects, $\sim 85\%$ are below the SDSS *i*-band flux limit, and so are more likely to be spurious sources. The remaining missing sources are likely due to astrometry changes, but these are an essentially negligible fraction in our final training set and catalogue (section 3.3).

The different epochs of the surveys utilized mean that observations span over 10 years in the observed frame (over 3 years at $z = 2$). Typical UV/optical variability over these rest-frame timescales is 0.1–0.2 mag, depending on luminosity (Hook et al. 1994). Variability in the IR over these timescales is also ~ 0.1 mag (*J*-band; Kozłowski et al. 2010). These variability levels are less than the average errors for individual filters, and definitely less than the range in colors, and are therefore largely unimportant for this work.

2.3 Relative-flux-redshift density model

XDQSOz models the relative-flux-redshift distribution with a large number of Gaussians by deconvolving the distribution for the training sets described above with the XD technique (Bovy, Hogg & Roweis 2011). This method is uniquely suited to the data available for modeling quasar flux space, because not all objects have all fluxes available and the uncertainties are heterogeneous. XD assumes that the flux uncertainties are Gaussian (as they are for the SDSS; Ivezić et al. 2003, 2007). The spectroscopic redshifts are assumed to have null uncertainties because the typical values are on the order of 10^{-3} (Schneider et al. 2010), far smaller than the errors on photometric redshifts.

As in Bovy et al. (2012), we use a sum of 60 Gaussians to model the deconvolved flux-redshift density. All fits are performed in 0.2 mag wide bins of *i*-band magnitude, starting at $i = 17.7$ up to $i = 22.5$ (for 47 overlapping bins centered 0.1 mag apart). All of the quasars are used in the fit to each bin, by rescaling the fluxes according to the luminosity function as described in section 2.2.1. The model contains a total of $47 \times (60 \times [1 + d + d(d+1)/2] - 1)$ parameters — for five SDSS fluxes, two GALEX fluxes, four UKIDSS fluxes, and two WISE fluxes, this is a total of 296,053 parameters (plus an additional 85,493 from the stellar fits from XDQSO with all of the available fluxes).

Figures 2 and 3 show the flux-flux (*W1* and *W2* versus SDSS *r*) and color-color (*r* – *W1* versus *W1* – *W2*) plots in one *i*-band bin ($18.6 \leq i < 18.8$) for a random re-sampling from the best-fit XD model (convolved with actual data errors, left columns) and the data these fits are based on, resampled according to the quasar luminosity function (right columns). Note that all of the WISE data are kept in the native Vega magnitude system. Figures 4 and 5 show similar plots comparing samplings of the redshift fits to the redshifts of the real data. The resampled error-convolved fits to the data are nearly identical to the real data in all cases, illustrating

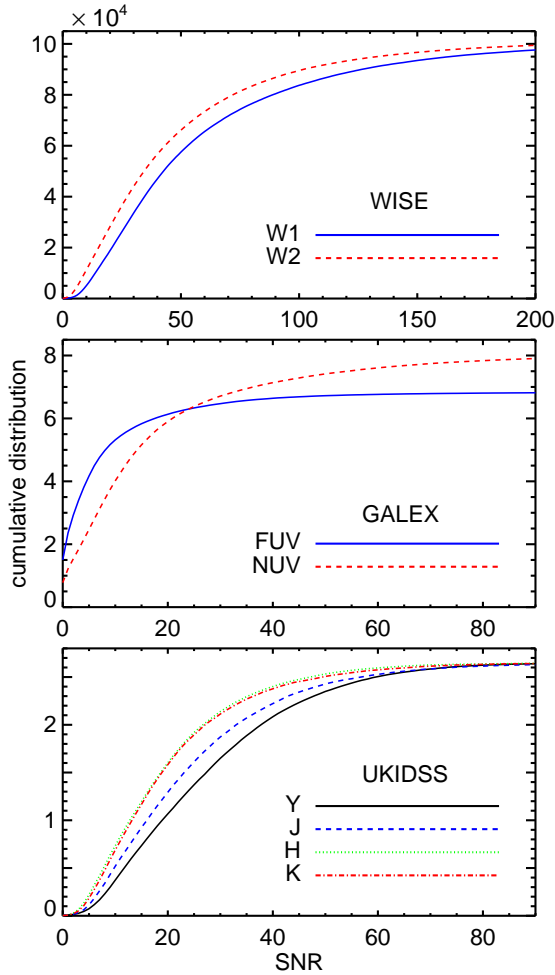


Figure 1. Cumulative distributions of the SNR for *WISE*, *GALEX*, and *UKIDSS* photometric data of spectroscopically confirmed quasars. *WISE* has data available for more sources, and is also deeper than the other surveys.

the accuracy of the fits. We also point out in Figure 5 that above $z \sim 3$ the colors fall below $W1 - W2 = 0.8$, a common cut used to select quasars from *WISE* data alone (e.g. Stern et al. 2012, see section 3.2)

3 RESULTS & DISCUSSION

3.1 Performance

Sections 5 and 6 of Bovy et al. (2012) already demonstrated that *XDQSOz* outperforms many other methods when calculating quasar probabilities and photometric redshifts. Here, we simply present the improvements with the addition of *WISE* flux information. In order to assess the performance of *XDQSOz*, we calculate the quasar probability P_{QSO} for the known SDSS DR7 spectroscopic quasars as well as the stars in our training set. We use the entire samples for all of the below analysis, including additional fluxes where they are available (and the analysis requires them).

First, we present the results using the broad redshift bins of the original *XDQSO* ($z < 2.2$, $2.2 < z < 3.5$, and $z > 3.5$), shown in Figures 6 (quasars) and 7 (stars). These figures show a comparison between using only SDSS fluxes, SDSS+*WISE*, and

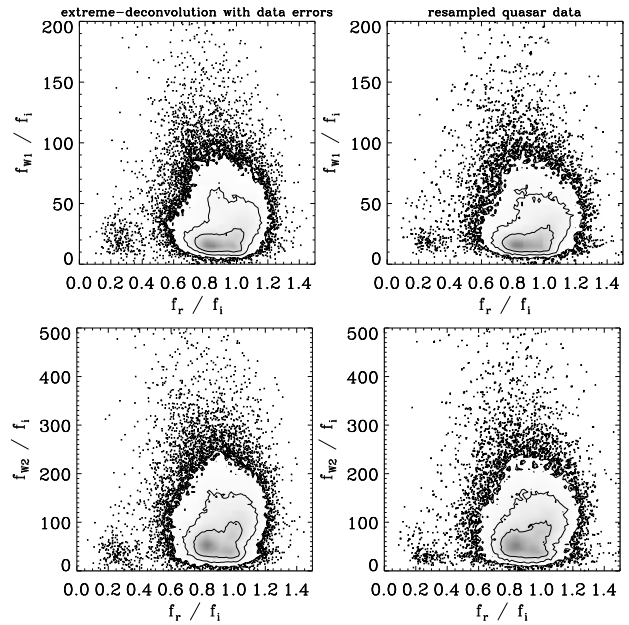


Figure 2. *W1* (top) and *W2* (bottom) versus SDSS r (all normalized by the i -band flux), illustrating a projection of the space in which the fits are performed. The left columns show a random sampling from the *XD* fits with errors from the real data added in, and the right columns show the real quasar data resampled according to the quasar luminosity function (see sections 2.2.1 and 2.3). Note that the *WISE* fluxes retain their native Vega zero-point, while the SDSS fluxes are in AB, hence the large relative *WISE*-SDSS fluxes.

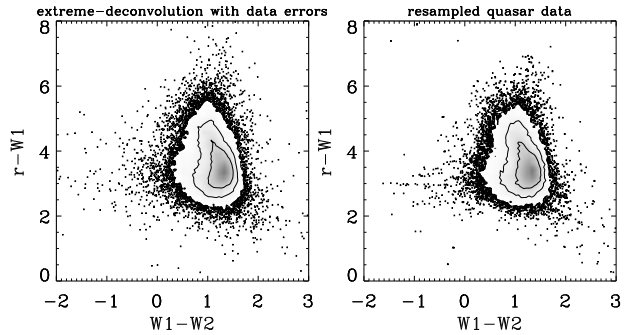


Figure 3. The same as Figure 2, but for $r - W1$ versus $W1 - W2$ colors.

SDSS+*GALEX*+*UKIDSS*+*WISE*. The percentages of the sub-samples belonging to the redshift bins of interest that are identified as quasars at $P > 0.5$ and $P > 0.8$ are given in each panel for easy comparison.

The method does an excellent job identifying quasars in the appropriate redshift bins, and not as stars, even with only the SDSS fluxes. However, there is significant improvement with the addition of *WISE* information, especially at $z > 2.2$. At low redshift, gains are on the order of $\sim 5\%$. In the mid-redshift range, the improvement is $\sim 15\text{--}20\%$. At high redshifts, the improvement is $\sim 10\text{--}15\%$. The addition of *GALEX* and *UKIDSS* data adds improvements of a few percent at most over the SDSS+*WISE* information — adding *WISE* fluxes has the single greatest effect on the performance of *XDQSOz* when classifying quasars. Similar results are found when testing the stellar training set (Figure 7). Very few stars are identified as quasars, with a significant improvement when *WISE* data are added and only a marginal further increase with *GALEX* and

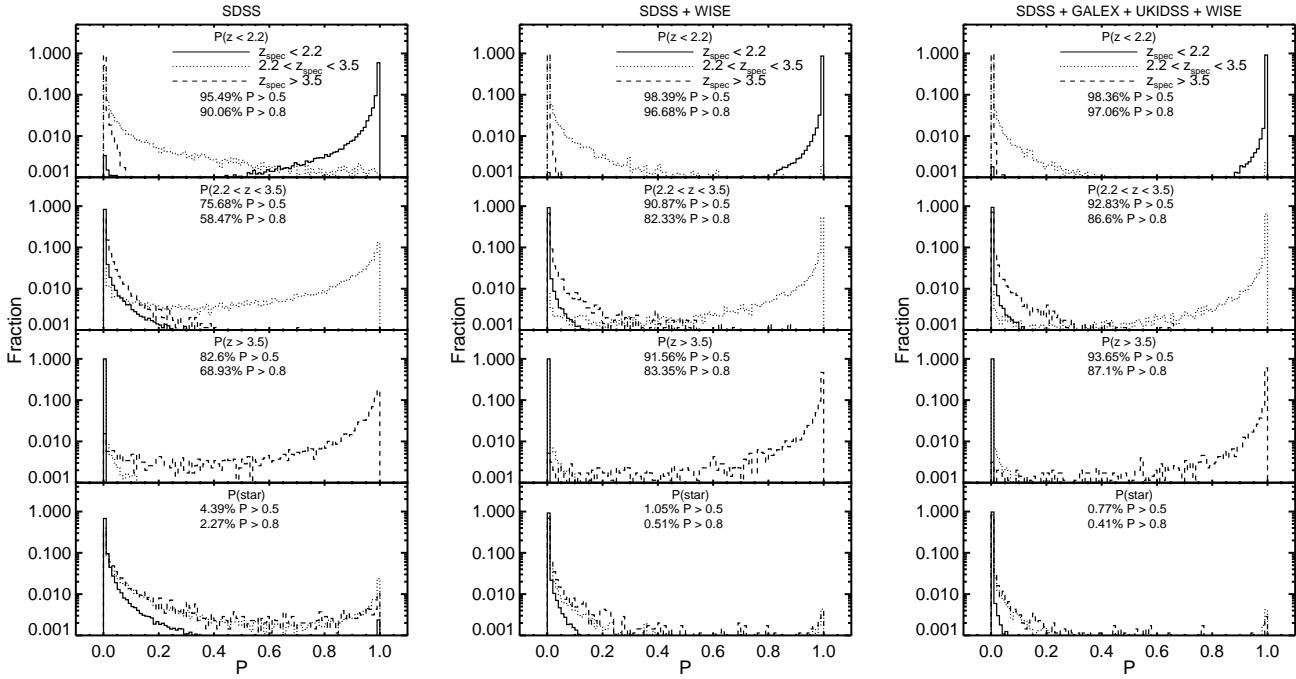


Figure 6. The XDQSOz probability that a spectroscopically confirmed quasar is a quasar in broad redshift bins, using just SDSS photometry (left), SDSS+WISE photometry (center), and SDSS+GALEX+UKIDSS+WISE photometry (right). Within each panel are the distributions of the probabilities that objects are low-redshift quasars ($z < 2.2$, first panel), medium-redshift quasars ($2.2 < z < 3.5$, second panel), high-redshift quasars ($z > 3.5$, third panel), or stars (bottom panel). Line styles indicate the spectroscopic classification of each subset. The percentage of objects that belong to each redshift bin of interest and meet the thresholds $P_{\text{QSO}} > 0.5$ and $P_{\text{QSO}} > 0.8$ are listed in each panel. In the case of P_{star} , these percentages are for quasars at all redshifts. Similar figures showing additional combinations of data are provided in the appendix.

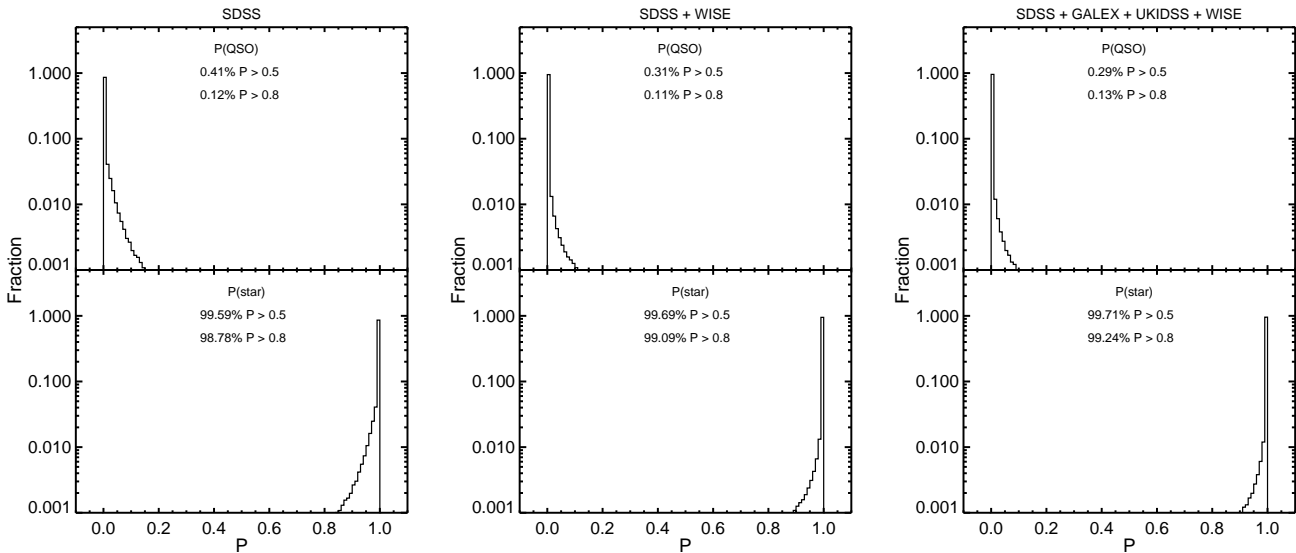


Figure 7. The XDQSOz probabilities for the stellar training set, using various combinations of photometric data. The top panels shows the probability that an object is a quasar at any redshift, and the bottom shows the probability that it is a star. Note that there is some contamination from quasars in the stellar training set (on the order of $\sim 1\%$). Similar figures showing additional combinations of data are provided in the appendix.

UKIDSS information. Similar plots for other combinations of data are provided in the appendix for comparison.

The top panels of Figure 8 show a similar analysis, using probabilities that an object is a quasar at *any* redshift. With just the SDSS data alone, $\sim 90\%$ of the quasars are recovered with $P_{\text{QSO}} > 0.8$. The addition of WISE photometry improves this dra-

atically, increasing the fraction with $P_{\text{QSO}} > 0.8$ to $\sim 97\%$. The further addition of GALEX and UKIDSS data only increases the performance by another $\sim 1\%$. Without the use of WISE these percentages only reach $\sim 93\%$ for any other combinations of data (see the appendices).

Because our training set has additional GALEX and UKIDSS

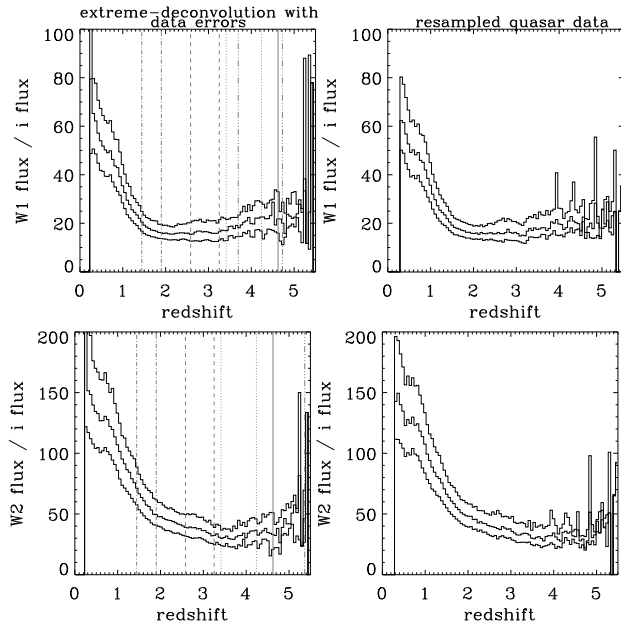


Figure 4. The same as Figure 2, but for $W1$ and $W2$ versus z , again showing a projection of the space in which fits are performed. The *WISE* fluxes retain their native Vega zero-point, hence the large relative *WISE*-SDSS fluxes. The vertical lines indicate the redshifts at which various strong quasar emission lines fall within any of the relevant filters (i , $W1$, or $W2$): solid is $\text{Ly}\alpha$, dotted is C IV , dashed is C III] , dash-dotted is Mg II , dash-dot-dot-dotted is $\text{H}\alpha$. None of these lines have a strong effect on the behavior of the normalized $W1$ or $W2$ fluxes as a function of redshift.

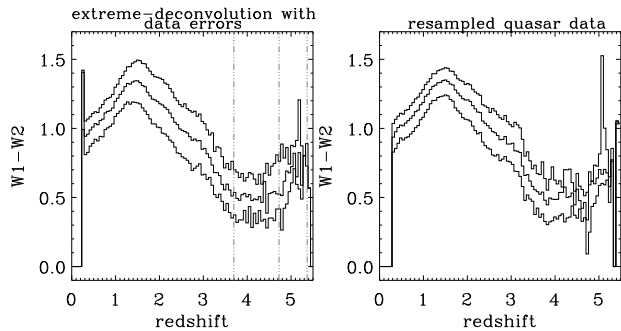


Figure 5. The same as Figure 4, but for $W1 - W2$ colors instead of fluxes. The vertical lines indicate the range of redshifts for which $\text{H}\alpha$ lies in the $W1$ filter bandpass.

data not available in the training set of Bovy et al. (2012), we explicitly compare the percentage of known quasars identified at $P_{\text{QSO}} > 0.8$ using our new fits using SDSS+*GALEX* data and SDSS+UKIDSS data and the same fits of Bovy et al. (2012) available from the previous *XDQSOz* release. The new data do not greatly alter the fits, but the changes are large enough such that the new fits improve the ability of *XDQSOz* to identify quasars by $\sim 0.3\%$ for SDSS+*GALEX*, and $\sim 0.1\%$ for SDSS+UKIDSS.

The bottom panels of Figure 8 show the comparison of the peaks of the redshift PDFs versus the spectroscopic redshifts. These panels include all objects in all of the panels, regardless of which data are available for a given source (e.g. an object is included in the panel using fits to all of the fluxes even if it doesn't have all of the fluxes available) or if its redshift PDF has multiple peaks (see below). There is a clear improvement over just SDSS data with the

addition of *WISE*, and again only a marginal further improvement by adding *GALEX* and UKIDSS data. The groups of points that are far off-axis and most strongly present in the SDSS-only panel are largely due to strong quasar emission lines (e.g. Mg II or C IV) moving between filters (see e.g. Figure 12 of Ball et al. 2008), and additional data at other wavelengths helps to reduce these affects.

To better illustrate the benefit of *WISE*, Figure 9 shows the difference between the peak photometric redshifts and the spectroscopic redshifts for several combinations of data. The use of *WISE* data clearly has the single largest effect, with smaller additional gains from including *GALEX* and UKIDSS. In the case of SDSS+*WISE*, $\sim 95\%$ of the objects have $\Delta z = |z_{\text{spec}} - z_{\text{phot}}| < 0.3$ and $\sim 76\%$ have $\Delta z < 0.1$. The large bumps around $|\Delta z| = 1.5$ in most combinations of data without *WISE* represent the off-axis clumps discussed above, and are largely removed with the additional *WISE* data because only one emission line ($\text{H}\alpha$) affects the *WISE* bandpasses, and only at higher redshifts ($z \gtrsim 3.5$; see e.g. Figure 4).

If we limit the analysis of the photometric redshifts to only objects with photometry available in all filters, and only those that have single-peaked² redshift PDFs, the performance is outstanding. This is shown in Figure 10. In this case, $\sim 98\%$ of the 21,929 objects have $\Delta z < 0.3$, and $\sim 84\%$ have $\Delta z < 0.1$.

Figure 11 shows the redshift PDFs for four known quasars with various combinations of data. These are the same objects that are shown in Figure 12 of Bovy et al. (2012). In general we see that the addition of *WISE* data has the power to significantly narrow the PDF around the known spectroscopic redshifts. However, it is not always the case that SDSS+*WISE* has a similar accuracy as when all of the fluxes are used. The addition of *GALEX* and UKIDSS data often has a significant effect on the photo- z estimates. The appendix includes some examples of the effect of *WISE* data on the redshift estimation of high-redshift ($z > 3$) quasars.

² A peak in the redshift distribution is defined as any continuous region above the uniform distribution between $0.3 < z < 5.5$

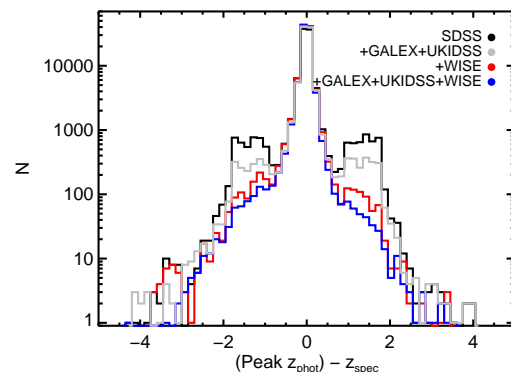


Figure 9. The distribution of the difference between the peak of the photometric z PDF and the spectroscopic z for the DR7 quasar sample, using several different combinations of data. Using *WISE* fluxes clearly narrows the distribution significantly, more so than using SDSS+*GALEX*+UKIDSS. Gains from using all of the available data are small compared to using just SDSS+*WISE*.

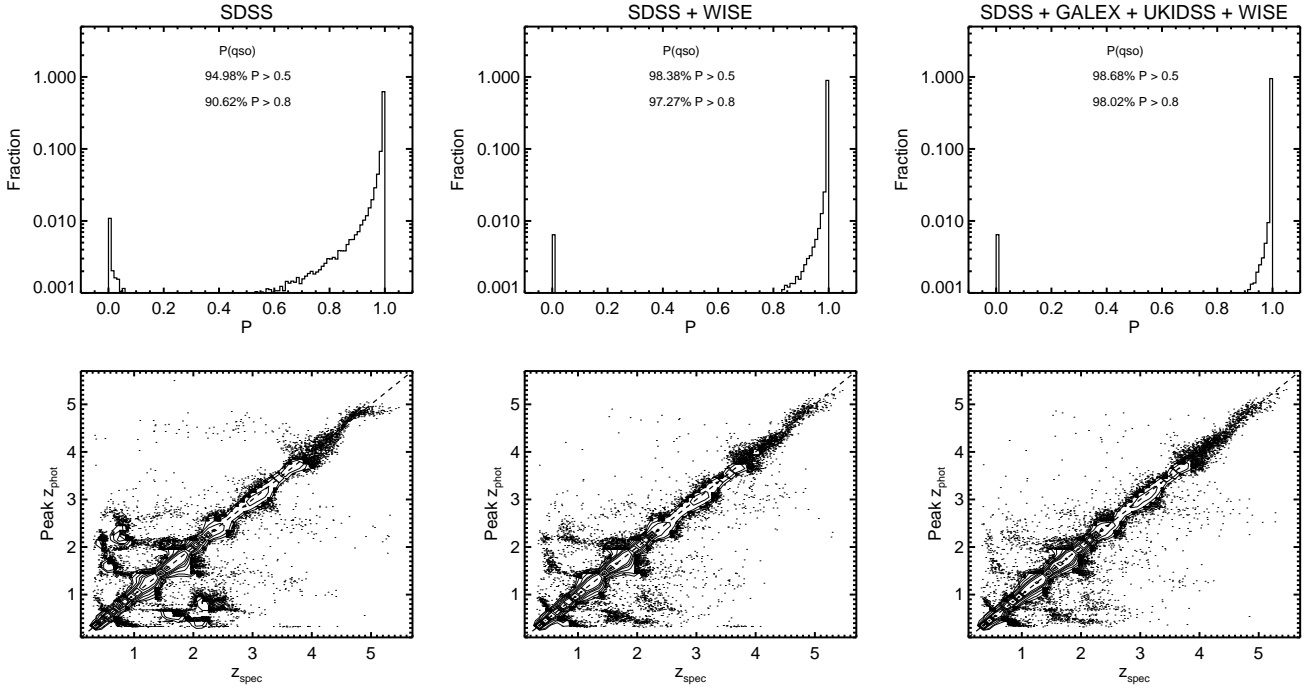


Figure 8. The three panels show the distributions of P_{QSO} over all redshifts for known spectroscopic quasars from SDSS DR7 (top panels), and a comparison of the spectroscopic redshifts and the peak of the photo- z PDFs (bottom panels). From left to right, the results using just SDSS, SDSS and WISE, and all available photometry. Similar figures showing additional combinations of data are provided in the appendix.

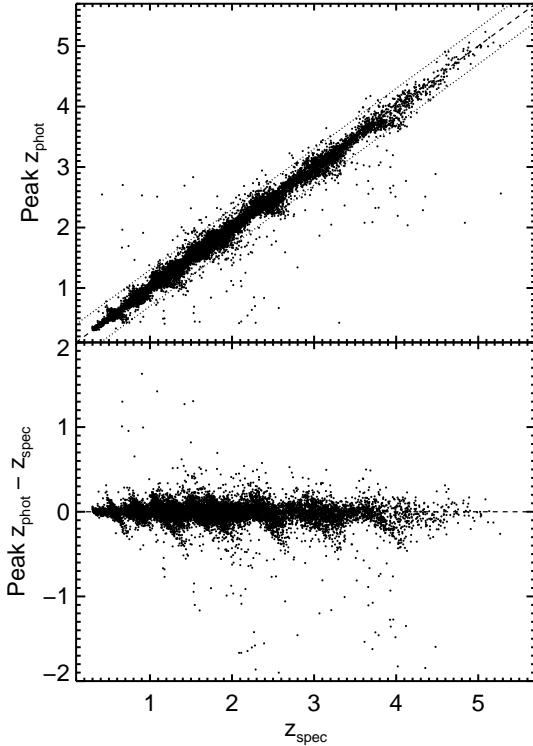


Figure 10. A comparison of the photometric versus spectroscopic redshifts, limited to only objects with all of the fluxes available and with single-peaked redshift PDFs. The dashed line shows the one-to-one line, and the dotted lines indicate $\Delta z = 0.3$.

3.2 Comparison with WISE color selection

How does the use of WISE photometry in the XDQSO method compare with pure WISE-color cuts to select AGN? Of course, with the current XDQSOz formalism (and the use of forced-photometered WISE data at SDSS positions) we are limited to objects with optical SDSS detections, and so we cannot directly compare the methods for selecting obscured objects that fall below the limits of SDSS due to obscuration of the quasar. However, it is fair to compare the two methods in identifying quasars with optical detections, as WISE colors are used to identify samples of unobscured quasars for comparison with obscured quasars (e.g. Donoso et al. 2014; DiPompeo et al. 2014, 2015). To examine this question, we start with a test sample of DR8 point sources and spectroscopically confirmed quasars from the DR7 and DR10 quasar catalogues (Schneider et al. 2010; Pâris et al. 2014) in a circular region centered at RA=180°, Dec=40° with a radius of 10°. The use of both quasar catalogues allows us to analyze the methods at high and low redshifts, as DR7 was effectively limited to $i < 20.2$ for high-redshift quasars, while the BOSS survey (included in DR10) specifically targeted $z > 2.2$ quasars with $g \leq 22$ OR $r \leq 21.85$. After applying the flag cuts used in the construction of the catalogue presented here (section 3.3) and limiting to sources with W1 and W2 forced photometry available, there are 904,571 point sources and 8,280 spectroscopically confirmed quasars in this region.

We analyze both the completeness ($N_{\text{QSO, sel}}/N_{\text{QSO}}$) and the selected target density ($N_{\text{point, sel}}/\text{area}$) for various cuts of P_{QSO} and W1 – W2 (always applying a cut at W2 < 15 in the latter case; Stern et al. 2012). Note again that all WISE magnitudes are in the Vega system. This analysis is performed separately for two redshift ranges, $z < 1$ and $z > 2.5$. The results are shown in Figure 12. We can see that the XDQSO method is far more complete than the simple WISE color cuts, especially for high redshift ob-

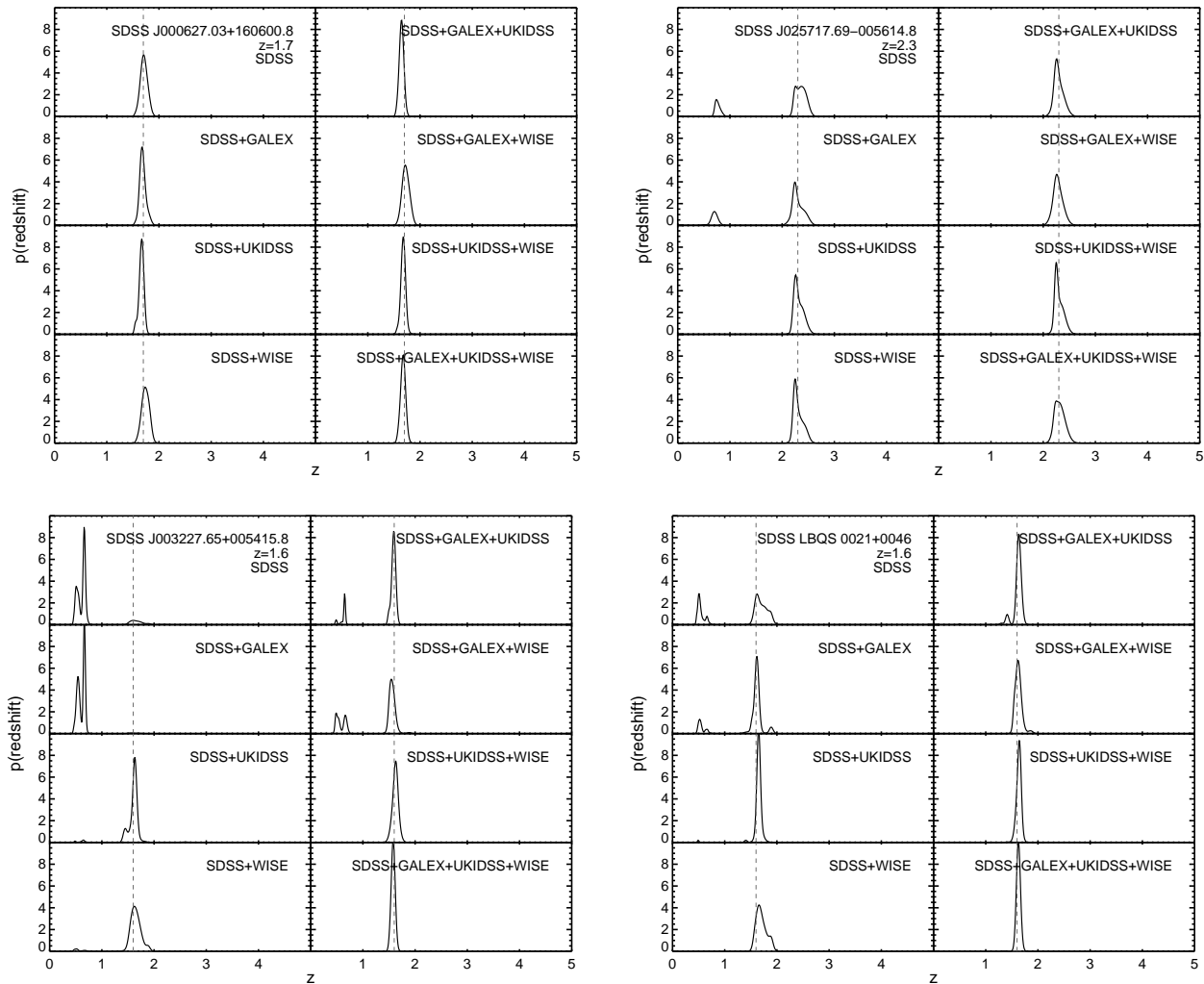


Figure 11. Examples of redshift PDFs for known spectroscopic quasars using various combinations of photometric data.

jects where the completeness is nearly an order of magnitude better. This is not particularly surprising, as it is well noted in the work of Stern et al. (2005), Stern et al. (2012), and Assef et al. (2013) that applying a simple color cut finds virtually no quasars at $z > 3$. Figure 13 shows that for SDSS spectroscopic quasars, this is largely due to the $W2$ limit — most SDSS quasars with $W2 > 15$ are at $z > 2$. This is at least partially because the higher z SDSS quasars are, on average, optically (and therefore probably bolometrically) fainter. However, Figure 5 also shows that most high- z objects have $W1 - W2$ colors that are too blue to make the cut. At lower redshift, using simple color cuts we find a completeness of $\sim 75\%$, in rough agreement with Stern et al. (2012). The most striking feature in these plots is that the most conservative cuts in $XDQSOz$ begin where the *WISE* color cuts plateau.

For a given target density³ the $XDQSO$ selection is more com-

³ We note that the target density at a cut of $W1 - W2 > 0.8$ is lower than what is found in Stern et al. (2012). This is because we require optical detections in the SDSS, which means that we are missing the most heavily obscured *WISE*-selected AGN. These objects may make up nearly half of the quasar population (Hickox et al. 2007; DiPompeo et al. 2014; Assef et al. 2014)

plete, indicating that this method is also more efficient. However, over all redshifts, to reach a completeness level above 90% requires target densities over 100 per deg².

3.3 Quasar catalogue and code

We release an updated version of the quasar catalogue of Bovy et al. (2011), with the additional use of *WISE* data in our fits to the flux-redshift density space. The catalogue includes quasar/star probabilities for all SDSS DR8 point sources (`objc_type` = 6) with a reasonable detection (extinction-corrected magnitude in at least one band is above the completeness limit; Aihara et al. 2011), a dereddened i -band magnitude in the range $17.75 \leq i < 22.45$, and with $P_{QSO} > 0.2$. A catalogue of quasar/star probabilities for *all* of the SDSS DR8 point sources is available upon request. The catalogue includes the entries listed in Table 2 of Bovy et al. (2011), along with 6 additions: `galex_matched`, `galex_used`, `ukidss_matched`, `ukidss_used`, `wise_matched`, and `wise_used`. The “matched” tags indicate whether an object was detected in forced-photometry of each survey (true or false), and

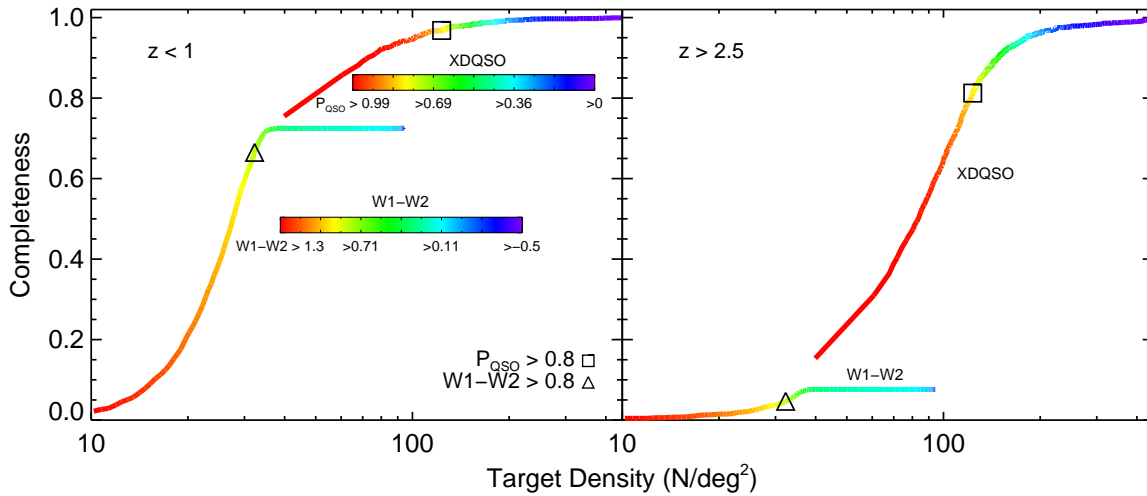


Figure 12. A comparison of the completeness and efficiency in identifying quasars as a function of *XDQSOz* quasar probabilities and simple *WISE* color cuts (combined with a *W2* cut), using a test $\sim 300 \text{ deg}^2$ region, and the DR7 and DR10 spectroscopic quasar catalogues. The left and right panels show low ($z < 1$) and high ($z > 2.5$) redshift ranges, respectively. Colors indicate various cuts in P_{QSO} and $W1 - W2$, as shown in the color bars. Cuts at $P_{\text{QSO}} = 0.8$, the inflection point in the total (all redshifts) quasar completeness for *XDQSOz*, are marked with a square, and the standard $W1 - W2$ cuts are indicated by a triangle. *XDQSOz* is more complete and more efficient, especially at high redshift. This is largely because of the *W2* restriction (Figure 13). Impressively, *XDQSOz* picks up where *WISE* color cuts plateau.

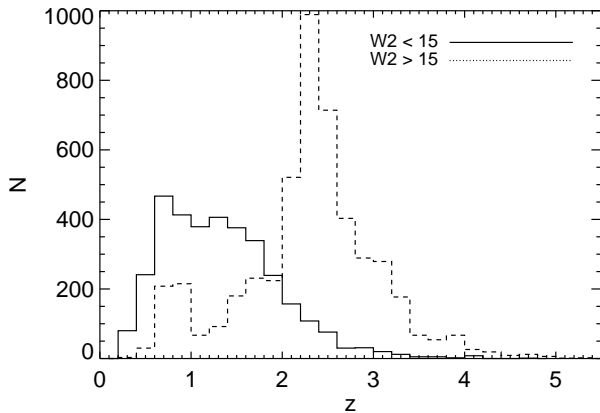


Figure 13. The distributions of redshift for SDSS quasars (from both the DR7 and DR10 quasar catalogues) with $W2 < 15$ (solid) and $W2 > 15$ (dashed). The $W2 < 15$ requirement for *WISE*-only quasar selection, which prevents contamination from massive, star-forming galaxies at high redshift, removes many high-redshift SDSS quasars as well. This at least partially reflects the fact that the SDSS quasar selection uses fainter limits at high redshift.

the “used” tags indicate if the given survey’s fluxes were used in the underlying model for the probability calculation.

This catalogue contains 5,537,436 (total; 3,874,639 weighted by probability) potential quasars, and the following additional tags of redshift information: *npeaks* (number of peaks in the redshift distribution, from 0 to 6), *peakz* (the redshift at the highest probability of the widest continuous region above the uniform distribution), *peakprob* (the probability associated with the peak redshift), *peakfwhm* (the FWHM of the primary peak), *otherz* (the redshifts of the peaks of up to six other regions above the uniform distribution), *otherprob* (the peak probabilities of up to six secondary peaks), and *otherfwhm* (the FWHM of up to six other secondary peaks).

The distribution of the peak photometric redshifts for this catalogue is shown in Figure 14. There is a peak around $z \sim 1$, where star/galaxy separation becomes difficult for faint objects (see below). Many of these sources are likely galaxies with quasar-like colors, and are potentially borderline objects. Somewhat surprising is the additional peak at $z \sim 5$, given that high redshift quasars are relatively rare. In visual inspection of the SDSS imaging around 100 random sources with photometric redshifts of ~ 5 , we find that approximately 20% have nearby bright red stars that likely contaminate the photometry. Indeed, the majority of these objects have the SDSS SUBTRACTED flag⁴ set. We extend the BOSS bright star mask⁵ (as in, e.g. White et al. 2011, 2012) across the whole SDSS region and apply it, which removes a significant number of these objects. An additional tag has been added to the catalogue, *bright_star*, to indicate objects that may suffer from contamination from nearby bright stars because they lie within the bright star mask.

There are other problematic regions in the SDSS that may cause inaccurate quasar probabilities or redshifts, including fields with poor photometry and regions in the North Galactic Cap with bad *u*-band data. These are indicated in the catalogue in the tags *bad_field* and *bad_u*, respectively. DiPompeo et al. (2014) built a mask that removes regions around flagged/contaminated data in the *WISE* catalogue, including bright *WISE* stars (which

⁴ This flag indicates that the wings of the PSF of a bright star have been subtracted, as described in Stoughton et al. (2002) and at https://www.sdss3.org/dr8/algorithms/flags_detail.php

⁵ The original mask covering the BOSS footprint can be found at http://data.sdss3.org/datamodel/files/BOSS_LSS_REDUX/reject_mask/MASK.html. Our extended version is provided as a MANGLE polygon file along with several useful *WISE* masks as presented in DiPompeo et al. (2014), at <http://faraday.uwo.edu/~admyers/wisemask2014/wisemask.html>. Note that there is a typo (missing “~”) in the URL of DiPompeo et al. (2014).

may or may not be bright in the SDSS). An additional tag, `wise_flagged`, is included in the catalogue to indicate objects that lie in these regions.

However, even excluding objects that fall in these various masks, there is still a spike in the photometric redshift distribution of the catalogue around $z \sim 5$. Another possibility we tested is that red galaxies that fail star-galaxy separation in the SDSS DR8 imaging (i.e. they are unresolved) could mimic high- z quasars and cause this abundance of objects. Because the “stellar” training set is from co-added Stripe 82 data, the average seeing is better than that of the full catalogue and these contaminants would likely not be included in the training. However, if this were the case, we should see the spike at $z \sim 5$ reduced for the objects in the catalogue with the best seeing — this is not the case. The distribution of peak- z values is not a function of the seeing. It is not immediately clear in the SDSS imaging that there are obvious problems with the vast majority of these $z \sim 5$ sources, but we recommend users exercise caution with these objects until proper follow-up is performed.

One way to limit some of the potential contamination from e.g. failing star-galaxy separation is to limit the sample to relatively bright ($g < 21.5$) objects. Indeed, as shown with the dotted histogram in Figure 14, doing so removes the spikes in redshift and thus likely provides a more pure quasar sample.

We stress that this catalogue is probabilistic in nature, so many of the objects are likely not quasars, and it is not intended to represent, on its own, a statistical sample. This is largely due to the heterogeneous nature of the SDSS data it is built from, and it is possible to compile a subset of the catalogue that represents a complete statistical quasar sample. This has been done with previous versions of the *XDQSOz* catalog for an array of studies, including: probing the intergalactic medium with close quasar pairs (e.g. Prochaska et al. 2013; Hennawi & Prochaska 2013; Prochaska, Lau & Hennawi 2014; Rubin et al. 2014), photometric clustering to probe primordial non-Gaussianity (Ho et al. 2013; Agarwal, Ho & Shandera 2014; Leistedt & Peiris 2014; Leistedt, Peiris & Roth 2014), probing the extent of the Galactic halo (e.g. Deason et al. 2014), and studying baryon acoustic oscillations at high redshift (e.g. Slosar et al. 2013). The improved catalog here is sure to be a boon to future work in these areas, and is also useful for cross-matching with other catalogues/wavelengths in order to estimate quasar likelihoods or redshifts, analyses that do not require statistically complete samples, and searches for unusual objects.

The probabilistic quasar catalogue is available as a fits file (easily downloaded with a `WGET` command) at http://www.mpia.de/homes/joe/xdqsozcat_galex_ukidss_wise_p20.fits.gz. The updated *XDQSO* and *XDQSOz* codes for target classification and photometric redshift estimation, as well as the new flux-redshift density models including *WISE* data, are publicly available as a GitHub repository at <https://github.com/xdqso/xdqso/>. The FITS files containing the model are identical to those in the previous release of *XDQSOz*, with additional dimensions added to the end of each Gaussian component containing *W1* and *W2* information (in that order).

4 CONCLUSION

We have presented an update to the *XDQSOz* method of Bovy et al. (2011, 2012) for quasar classification and photometric redshift estimation that incorporates the two most sensitive *WISE* bands (3.6 and 4.6 μm , or *W1* and *W2*, respectively) into the relative-flux-

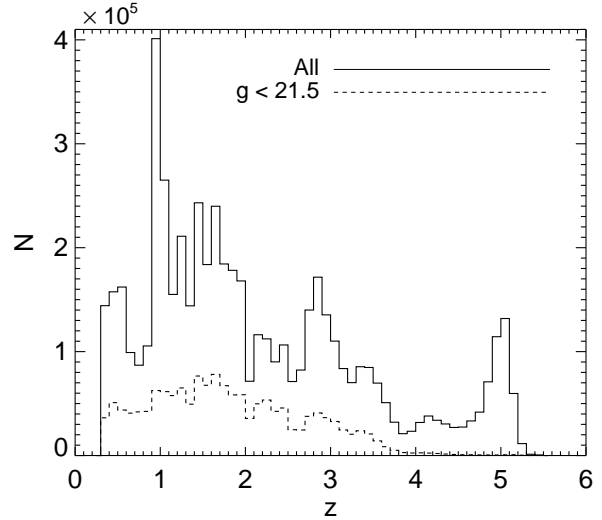


Figure 14. The peak photometric redshift distribution for the new quasar catalogue (solid line). The peak around $z = 5$ can be reduced by applying additional cuts to the data to remove artifacts or regions of bad photometry (see section 3.3), but many of the sources appear real in optical images. Additional follow-up is needed to determine if they are truly high- z quasars. Both this peak and the one at $z \sim 1$ are removed by limiting to relatively optically bright objects ($g < 21.5$; dotted line), which likely removes faint, possibly borderline quasars that fail star-galaxy separation.

redshift density model. The use of *WISE* information greatly enhances the precision of the method in identifying quasars (by ~ 5 –20%, depending on the redshift range of interest), especially at $z > 2$. It also improves the overall accuracy of the photometric redshift estimation. This method has better completeness compared to simple *WISE* color cuts (for quasars with lower levels of obscuration such that they are still detected and unresolved by the SDSS), again especially at high redshift, at similar efficiency. We present a catalogue of potential quasar candidates ($P_{\text{QSO}} \geq 0.2$) with photometric redshift estimates. This catalogue can be a powerful tool for identifying quasar samples or estimating redshifts for a wide variety of studies, and the improved accuracy will benefit future large quasar surveys.

ACKNOWLEDGEMENT

MAD and ADM were partially supported by NASA through ADAP award NNX12AE38G and EPSCoR award NNX11AM18A and by the National Science Foundation through grant numbers 1211096 and 1211112. J.B. was supported by NASA through Hubble Fellowship grant HST-HF-51285.01 from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. We thank David Hogg for his feedback on the manuscript and his work on the original *XDQSOz*.

REFERENCES

- Abazajian K. N. et al., 2009, *ApJS*, 182, 543
- Agarwal N., Ho S., Shandera S., 2014, *JCAP*, 02, 038
- Ahn C. P. et al., 2012, *ApJS*, 203, 21
- Aihara H. et al., 2011, *ApJS*, 193, 29
- Assef R. J. et al., 2014, *arXiv*, 1092
- Assef R. J. et al., 2013, *ApJ*, 772, 26

Baldwin J. A., 1977, *ApJ*, 214, 679
 Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tchong D., 2008, *ApJ*, 683, 12
 Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tchong D., Llorà X., 2007, *ApJ*, 663, 774
 Bovy J. et al., 2011, *ApJ*, 729, 141
 Bovy J., Hogg D. W., Roweis S. T., 2011, *Ann. App. Stat.*, 5, 1657
 Bovy J. et al., 2012, *ApJ*, 749, 41
 Budavári T. et al., 2001, *AJ*, 122, 1163
 Dawson K. S. et al., 2013, *AJ*, 145, 10
 Deason A. J., Belokurov V., Koposov S. E., Rockosi C. M., 2014, *ApJ*, 787, 30
 DiPompeo M. A., Myers A. D., Hickox R. C., Geach J. E., Hainline K. N., 2014, *MNRAS*, 442, 3443
 DiPompeo M. A., Myers A. D., Hickox R. C., Geach J. E., Holder G., Hainline K. N., Hall S. W., 2015, *MNRAS*, 446, 3492
 Donoso E., Yan L., Stern D., Assef R. J., 2014, *ApJ*, 789, 44
 Eisenstein D. J. et al., 2011, *AJ*, 142, 72
 Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *Astronomical Journal* v.111, 111, 1748
 Giannantonio T. et al., 2006, *PhRvD*, 74, 63520
 Giannantonio T., Scranton R., Crittenden R. G., Nichol R. C., Boughn S. P., Myers A. D., Richards G. T., 2008, *PhRvD*, 77, 123520
 Hennawi J. F., Prochaska J. X., 2013, *ApJ*, 766, 58
 Hickox R. C. et al., 2007, *ApJ*, 671, 1365
 Hickox R. C. et al., 2011, *ApJ*, 731, 117
 Ho S. et al., 2013, *arXiv*, 2597
 Hook I. M., McMahon R. G., Boyle B. J., Irwin M. J., 1994, *MNRAS*, 268, 305
 Hopkins P. F., Richards G. T., Hernquist L., 2007, *ApJ*, 654, 731
 Ivezić Ž. et al., 2003, *Mem. Soc. Astron. Ital.*, 74, 978
 Ivezić Ž. et al., 2007, *AJ*, 134, 973
 Kirkpatrick J. A., Schlegel D. J., Ross N. P., Myers A. D., Hennawi J. F., Sheldon E. S., Schneider D. P., Weaver B. A., 2011, *ApJ*, 743, 125
 Kozłowski S. et al., 2010, *ApJ*, 716, 530
 Lang D., 2014, *AJ*, 147, 108
 Lang D., Hogg D. W., Schlegel D. J., 2014, *arXiv*, 1410.7397
 Leistedt B., Peiris H. V., 2014, *MNRAS*, 444, 2
 Leistedt B., Peiris H. V., Roth N., 2014, *Physical Review Letters*, 113, 221301
 Martin D. C. et al., 2005, *ApJ*, 619, L1
 Mateos S., Alonso-Herrero A., Carrera F. J., Blain A., Severgnini P., Caccianiga A., Ruiz A., 2013, *MNRAS*, 434, 941
 Mateos S. et al., 2012, *MNRAS*, 426, 3271
 Myers A. D., Brunner R. J., Nichol R. C., Richards G. T., Schneider D. P., Bahcall N. A., 2007a, *ApJ*, 658, 85
 Myers A. D., Brunner R. J., Richards G. T., Nichol R. C., Schneider D. P., Bahcall N. A., 2007b, *ApJ*, 658, 99
 Myers A. D. et al., 2006, *ApJ*, 638, 622
 Pâris I. et al., 2014, *A&A*, 563, 54
 Prochaska J. X. et al., 2013, *ApJ*, 776, 136
 Prochaska J. X., Lau M. W., Hennawi J. F., 2014, *ApJ*, 796, 140
 Richards G. T. et al., 2009a, *AJ*, 137, 3884
 Richards G. T. et al., 2001a, *AJ*, 121, 2308
 Richards G. T. et al., 2009b, *ApJS*, 180, 67
 Richards G. T. et al., 2004, *ApJS*, 155, 257
 Richards G. T. et al., 2001b, *AJ*, 122, 1151
 Ross N. P. et al., 2012, *ApJS*, 199, 3
 Rubin K. H. R., Hennawi J. F., Prochaska J. X., Simcoe R. A., Myers A., Lau M. W., 2014, *arXiv*

Schneider D. P. et al., 2010, *AJ*, 139, 2360
 Scranton R. et al., 2005, *ApJ*, 633, 589
 Slosar A. et al., 2013, *JCAP*, 04, 026
 Stern D. et al., 2012, *ApJ*, 753, 30
 Stern D. et al., 2005, *ApJ*, 631, 163
 Stoughton C. et al., 2002, *AJ*, 123, 485
 Suchkov A. A., Hanisch R. J., Margon B., 2005, *AJ*, 130, 2439
 White M. et al., 2011, *ApJ*, 728, 126
 White M. et al., 2012, *MNRAS*, 424, 933
 Wright E. L. et al., 2010, *AJ*, 140, 1868
 Yip C. W. et al., 2004, *AJ*, 128, 2603
 York D. G. et al., 2000, *AJ*, 120, 1579

APPENDIX A: ADDITIONAL COMBINATIONS OF DATA

Given that this work uses updated catalogues of *GALEX* and UKIDSS forced-photometry, in addition to the new *WISE* data, we have performed fits to the relative-flux-redshift space using all possible combinations of data so that the appropriate model can be used for any combination of available data. These are included in the new release of the *XDQSOz* code. Figures A1, A2, and A3 show the distributions of probabilities returned for the known training sets and the photometric versus spectroscopic redshifts using these various combinations of data, for comparison with Figures 6, 7, and 8.

APPENDIX B: HIGH-REDSHIFT QUASARS

One of the motivations for incorporating *WISE* data into the *XDQSO* training is because of the utility of the mid-IR in finding highly reddened and/or high-redshift quasars that are missed using pure optical selection. Here we illustrate the improvement in constraining the redshifts of high- z quasars when *WISE* fluxes are incorporated. Because these are examples using known spectroscopic quasars, they are fairly bright in the optical — the use of *WISE* data will likely show an even more dramatic improvement in identifying these objects when the optical data is fainter.

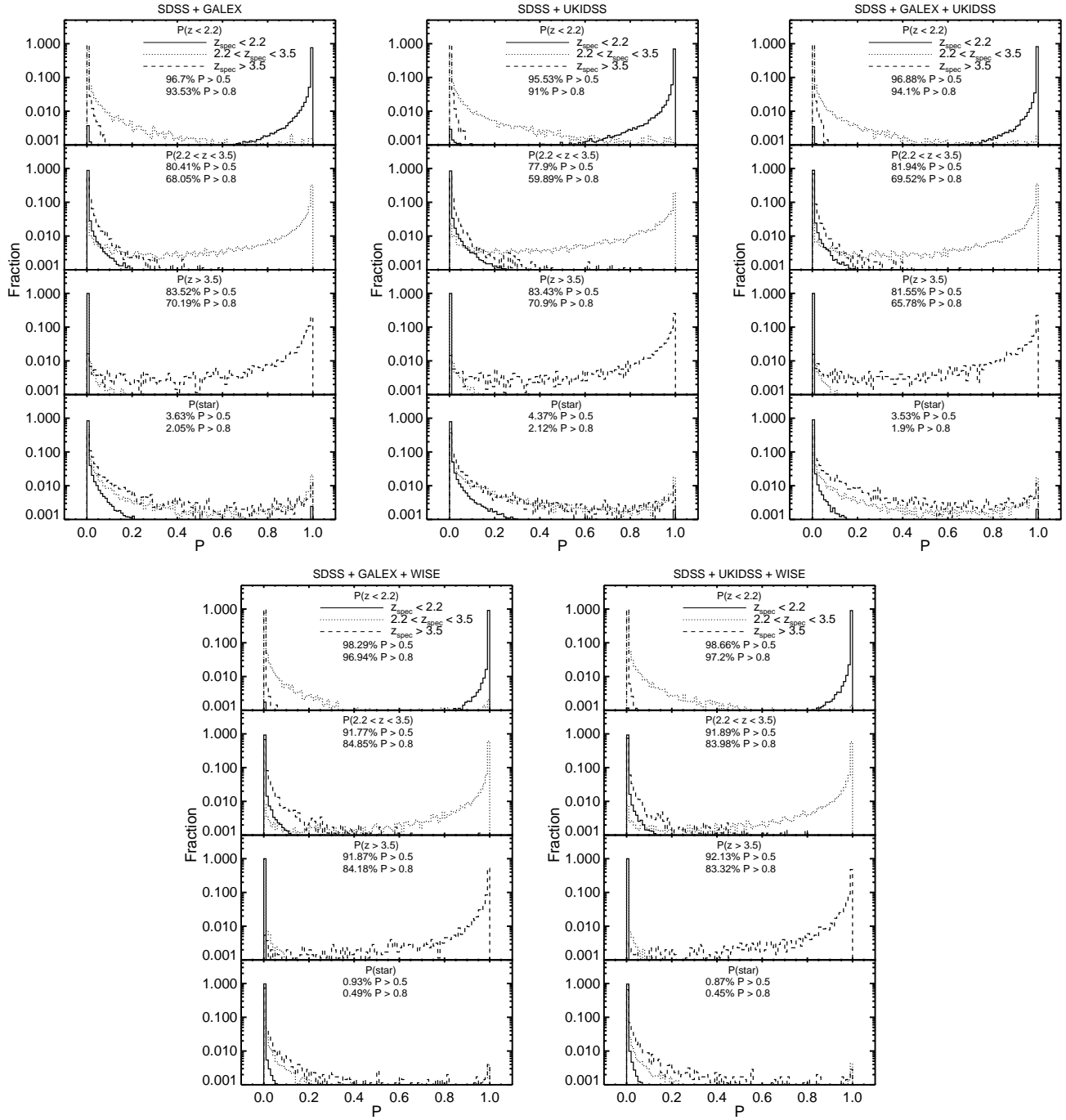


Figure A1. The same as Figure 6, illustrating the performance of $XDQSOz$ on known quasars in broad redshift bins but for other combinations of photometric data.

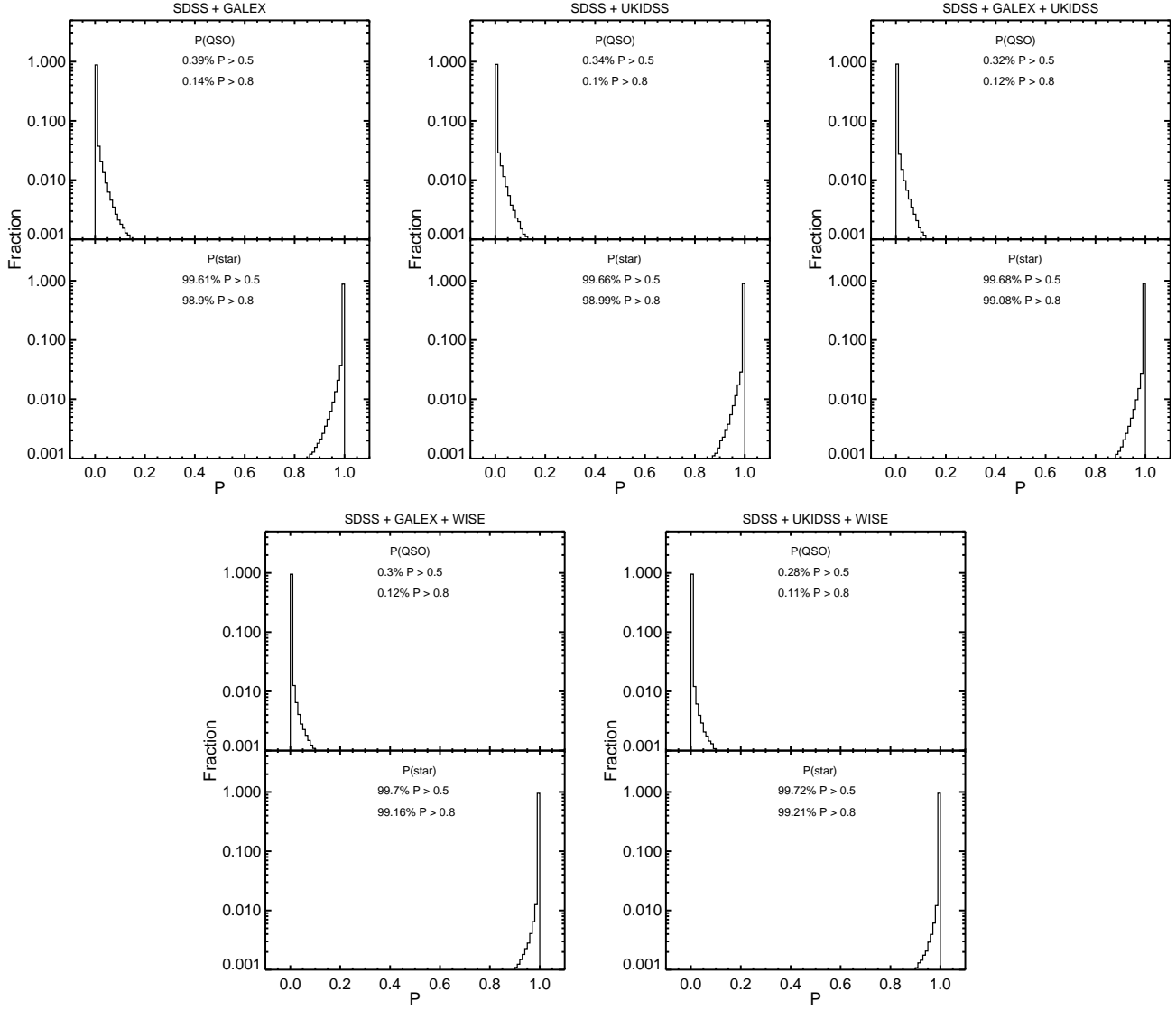


Figure A2. The same as Figure 7, illustrating the performance of XDQSOz on known stars but for other combinations of photometric data.

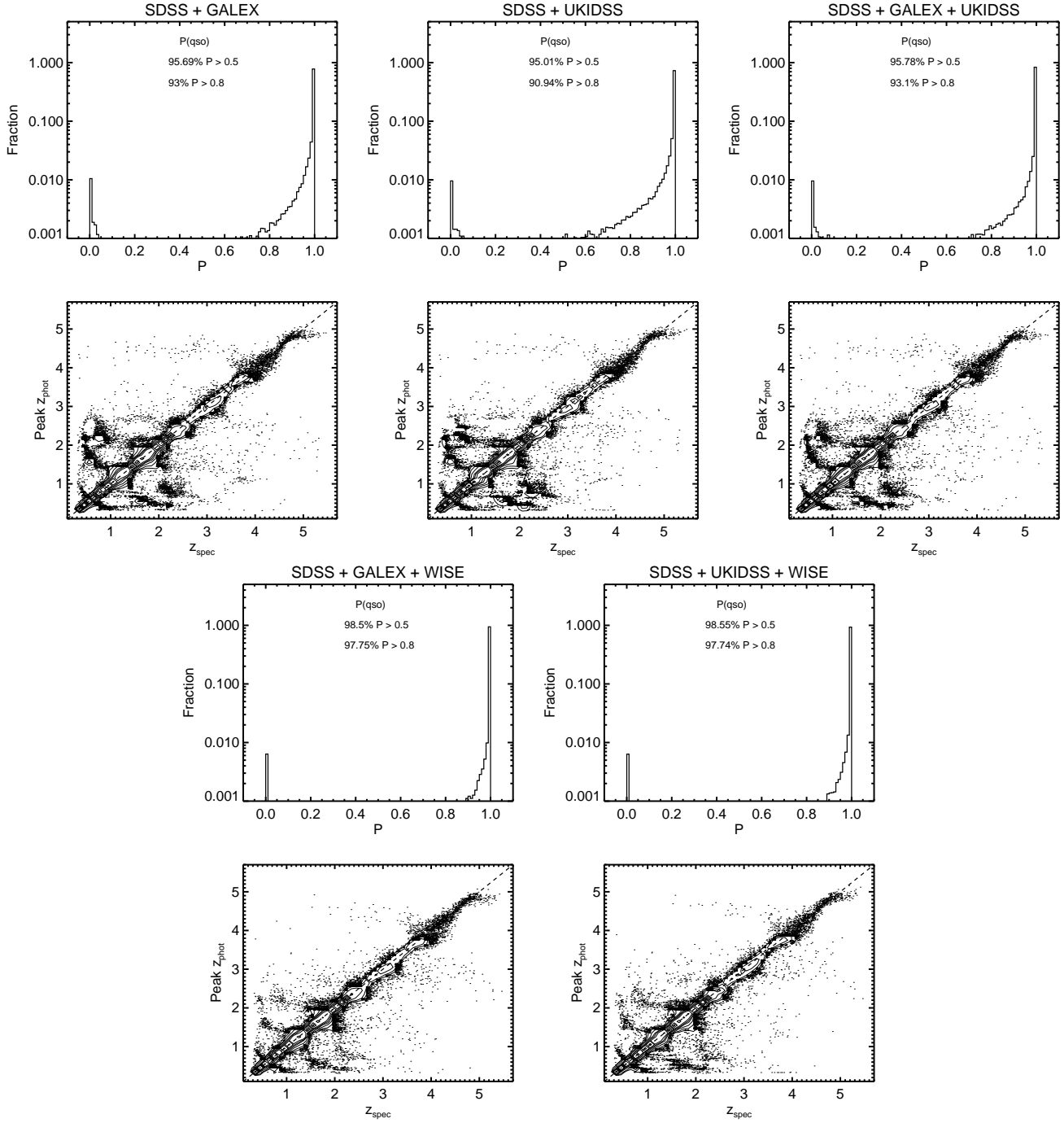


Figure A3. The same as Figure 8, illustrating the performance of $XDQSOz$ on known quasars at any redshift but for other combinations of photometric data.

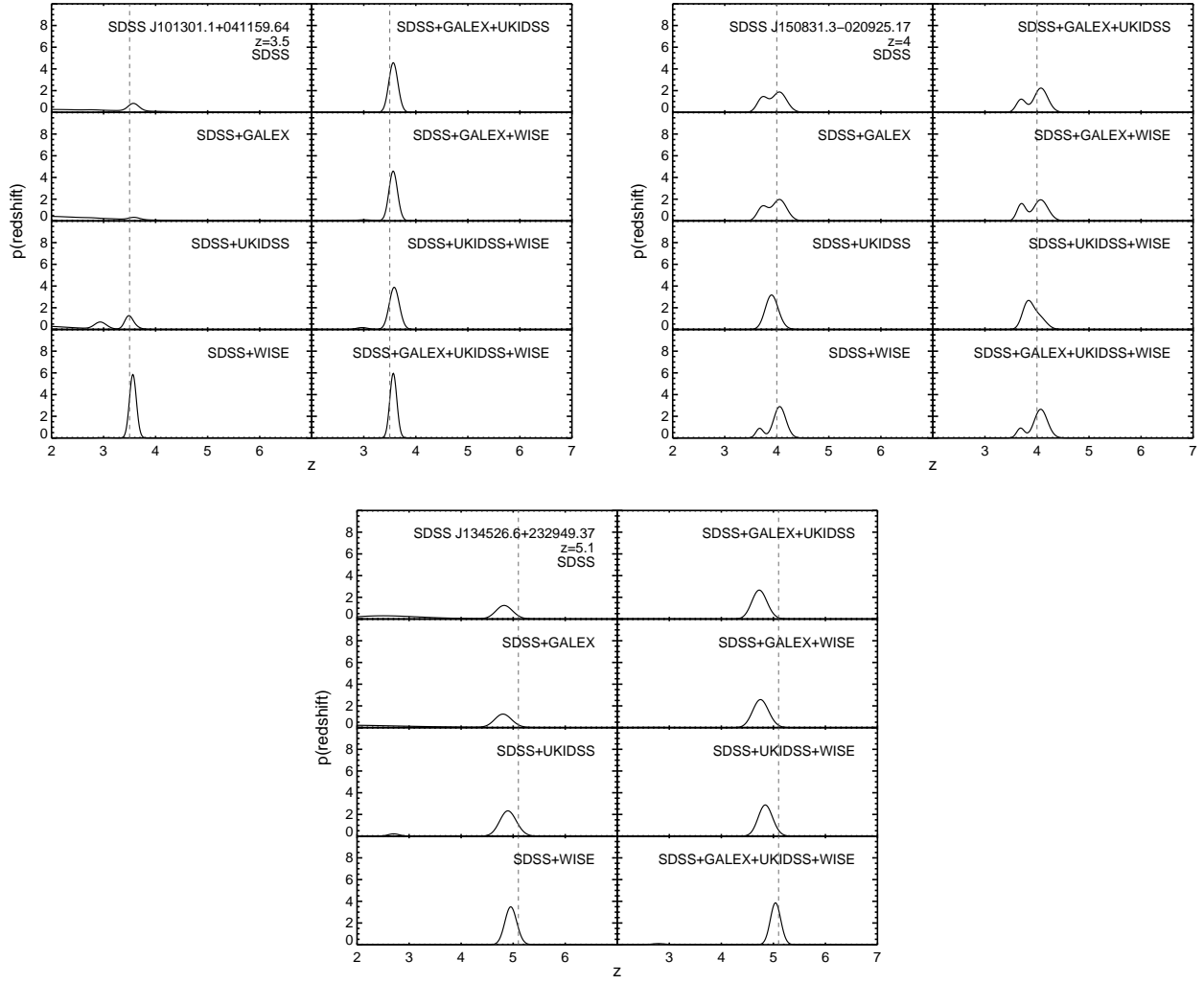


Figure B1. Examples of the effect of *WISE* information on the photometric redshift estimation for quasars at high redshift ($z = 3.5$, top left; $z = 4$, top right; $z = 5$, bottom).